

Research article

Open Access

Bioinformatic screening of human ESTs for differentially expressed genes in normal and tumor tissues

Abdel Aouacheria*^{†1,3}, Vincent Navratil^{†1}, Audrey Barthelaix²,
Dominique Mouchiroud¹ and Christian Gautier¹

Address: ¹Laboratoire de Biométrie et Biologie Evolutive, CNRS UMR 5558, Université Claude Bernard Lyon 1, 69622 Villeurbanne Cedex, France, ²Aptanomics, 181-203, avenue Jean Jaurès 69007 Lyon, France and ³Current address: Apoptosis and Oncogenesis Laboratory, IBCP, UMR 5086 CNRS-UCBL, IFR 128, Lyon, France

Email: Abdel Aouacheria* - a.aouacheria@ibcp.fr; Vincent Navratil - navratil@biomserv.univ-lyon1.fr;
Audrey Barthelaix - audrey.barthelaix@aptanomics.com; Dominique Mouchiroud - mouchi@biomserv.univ-lyon1.fr;
Christian Gautier - cgautier@biomserv.univ-lyon1.fr

* Corresponding author †Equal contributors

Published: 26 April 2006

Received: 14 October 2005

BMC Genomics 2006, 7:94 doi:10.1186/1471-2164-7-94

Accepted: 26 April 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/94>

© 2006 Aouacheria et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Owing to the explosion of information generated by human genomics, analysis of publicly available databases can help identify potential candidate genes relevant to the cancerous phenotype. The aim of this study was to scan for such genes by whole-genome *in silico* subtraction using Expressed Sequence Tag (EST) data.

Methods: Genes differentially expressed in normal versus tumor tissues were identified using a computer-based differential display strategy. Bcl-xL, an anti-apoptotic member of the Bcl-2 family, was selected for confirmation by western blot analysis.

Results: Our genome-wide expression analysis identified a set of genes whose differential expression may be attributed to the genetic alterations associated with tumor formation and malignant growth. We propose complete lists of genes that may serve as targets for projects seeking novel candidates for cancer diagnosis and therapy. Our validation result showed increased protein levels of Bcl-xL in two different liver cancer specimens compared to normal liver. Notably, our EST-based data mining procedure indicated that most of the changes in gene expression observed in cancer cells corresponded to gene inactivation patterns. Chromosomes and chromosomal regions most frequently associated with aberrant expression changes in cancer libraries were also determined.

Conclusion: Through the description of several candidates (including genes encoding extracellular matrix and ribosomal components, cytoskeletal proteins, apoptotic regulators, and novel tissue-specific biomarkers), our study illustrates the utility of *in silico* transcriptomics to identify tumor cell signatures, tumor-related genes and chromosomal regions frequently associated with aberrant expression in cancer.

Background

Large-scale transcriptome analysis of genes that are differentially expressed in tumor tissues compared to their normal counterparts is an important route to the identification of candidates that could play a role in human malignancies. A number of techniques, ranging from differential display and nucleic acid subtraction to serial analysis of gene expression, expression microarrays and gene chips, have been used to the discovery of such aberrantly expressed cancer-related genes [1]. The well-established differential screening technology, that allows for the simultaneous comparison of multiple gene expression levels between two samples differing in tissue type and pathological state, has been the more extensively applied. This simple and powerful method could be performed either experimentally or, since late 1999, digitally using expression databases. The *computer-based* differential display methodology, also referred to as '*in silico* subtraction' or 'electronic northern' [2-7], could identify transcripts preferentially expressed or repressed in the tumor context by comparing cancerous libraries (present in publicly available databases) against the remaining libraries. Strikingly, only few attempts were made to apply *in silico* transcriptomics to genome-wide and multi-tissue screening of cancer genes [8-10]. Thus, given the continuous expansion of the EST databases, both in terms of sequence and source diversity, updated and independent transcriptomic analyses are permanently needed.

In this study, we mined EST libraries for genes differentially expressed in normal and tumor tissues by using a novel computational approach, with the assumption that both the up- and down-regulated pools might contain genes involved in tumorigenesis. This strategy identified differential expression profiles and cancer candidate genes which may be useful in future cancer research. Higher expression of the anti-apoptotic protein Bcl-xL in liver cancer specimens compared to normal liver was confirmed by immunoblot analysis. Strikingly, we found that most cancer-associated changes in gene expression corresponded to genes that were actually downregulated or repressed. The chromosomes and chromosomal regions most frequently associated with aberrant expression changes in tumor versus normal cells were also determined. This analysis suggests that, although genes differentially expressed in cancerous libraries are distributed throughout the genome, chromosomal 'hot spots' of candidate genes could be identified.

Results

Identification of differentially expressed genes between normal and cancer tissues

Genes differentially expressed in tumor libraries compared to their normal counterparts are likely to play important roles in cancer etiology or could constitute rel-

evant genetic markers for cancer diagnosis. Here, we have performed *in silico* differential display to identify novel and known cancer-associated genes by comparing all the libraries representing tumors to the corresponding normal libraries for each tissue type. Details about the data mining procedures are presented in Table 1. In order to be able to compare expression levels between normal and tumor state, we compared EST counts from non-normalized, non-subtracted cDNA libraries. To overcorrect for the false positive rate, we decided to perform the highly conservative Bonferroni correction. Using this procedure, a total of 673 genes showed differential expression in tumor versus normal libraries by a factor of 10 or higher (Additional File 1: 'Upregulated candidates complete list', and Additional File 2: 'Downregulated candidates complete list'), with about one third being up-regulated (299) and the remaining being down-regulated (539). The *in silico* subtraction also resulted in the identification of 181 and 336 genes predicted to be present or absent in the tumor types compared to normal tissues, respectively. Because these EST clusters were identified either in normal or tumor libraries, it was not possible to derive their expression ratio, so we decided to present them as separated tables (Additional File 3: 'Tumor specific candidates complete list', and Additional File 4: 'Normal specific candidates complete list'). However, these two groups of genes have been fused to the 'up-regulated' and 'down-regulated' pools in the subsequent analyses. All in all, a sum of 112 novel transcripts was also found (i.e. sequences for which no description was available at the time of the study). Noteworthy, *in silico* subtraction identified 14.5 % (154/1060) previously studied genes involved in oncogenesis, based on a list of ~ 2500 genes compiled as previously described [11]. Since the fraction of such reference genes in our initial data set was 7.5 % (2401/31800), our data mining protocol expectedly lead to a significant enrichment in cancer genes (p value = $2.2 \cdot 10^{-16}$; exact Fisher test). These previously characterized and well-studied genes include the *p57^{KIP2}* and *p19^{INK4d}* cyclin inhibitors, and the *ras-GAP*, *c-fos*, *ret* and *myc* oncogenes. Last, in order to independently verify the validity of the EST-based tissue profiles, SAGE data were used to give

Table 1: Overview of the EST-based data mining strategy. Screening for differentially expressed genes between normal and cancer tissues. EST counts in each analytical step. Total number of EST clusters in each class (upregulated, downregulated, tumor-specific or absent in tumors) was determined after Bonferroni corrected exact Fisher test.

Total RNA (Ensembl)	~ 31,800
Total ESTs (after clustering)	~ $3.3 \cdot 10^6$
Total clusters	26,601
Total up-regulated EST clusters	227
Total down-regulated EST clusters	473
Total EST clusters specific to tumors	173
Total EST clusters absent in tumors	308

an indication of the tissue distribution of our transcripts in normal tissues. While SAGE results specified by tissue type converged with the analysis of ESTs for 65,4 % (197/301), 53,9 % (91/169), and 53,2 % (93/171) of the examined hits in the 'down', 'up' and 'normal-specific' groups respectively, i.e. precisely in the classes where an expression in the normal condition was expected, this percentage decreased to 37,7 % (46/122) for the 'tumor-specific' group of transcripts.

Cancer candidate gene analysis

The first general observation that could be made from our results is that a same gene could be either up-regulated or repressed according to the tumor cell type, allowing identification of tissue-specific gene expression profiles in tumor versus normal cells. For instance, among the set of candidates with differential expression in cancer, we observed a massive down-regulation of several collagen alpha chain genes (but not beta chain genes) in various tumor tissues, including decreased expression of collagen alpha 2(I) (also termed *col1A2*) in skin, placenta, testis, eye and bone (see Figure 1). Interestingly, *col1A2* has been reported as a tumor suppressor gene that could inhibit *ras*-induced oncogenic transformation [12,13]. Apart from collagens, other types of proteins that could be used as useful biomarkers include cytokeratins (CK). CK are particularly interesting epithelia specific intermediate filaments because their degradation gives rise to soluble fragments, measurable in the blood of patients and capable of cancer monitoring [14]. Our results show that a total of 13 CK genes were differentially expressed between normal and malignant cells in 9 different tissues (Figure 1), allowing tissue-specific expression profiling (e.g. specific expression of CK 5, 13 and 16 in tumor brain). Additionally, in line with previous microarray data [15], we found that hair-specific type II keratin was overexpressed in breast tumors compared to normal breast. We further determined that over the 190 genes which displayed aberrant expression in more than one tissue, 131 were "deregulated" in the same way (either up- or down, Figure 2 and Additional File 5: 'Consistent candidates in multiple tissues'). Included in this list of 'consistent' candidates are 13 transcripts encoding different ribosomal components, in accordance with the increasing body of evidence from the literature that correlates changes in the protein synthesis machinery with cancer [16-18]. Specific signatures for ribosomal genes could be determined, e.g. downregulation of the genes encoding 60S ribosomal L37, L38 and L44 in libraries prepared from tumor skin and tumor blood, whereas placental cancer libraries appear to be specifically enriched in transcripts encoding 40S ribosomal S2, S3 and S17. As depicted in Figure 3 (for the full data set, see Additional File 6: 'Tissue specific candidates'), some genes display a tissue-specific pattern of differential expression in tumor types, thus making them candidates

for specific diagnostic markers. Among these 114 genes differentially expressed in only one tissue are *14-3-3 sigma* in brain tumors and *Bnip3L* in blood. This latter gene, belonging to the Bcl-2 family of apoptotic regulators, has been described as a potential tumor suppressor [19,20]. Last, it is worth noting that a novel member of the methyltransferase enzyme family (ENST00000270172), that contains clear transcriptional repressors [21,22], was found to be specifically overexpressed in placental tumors.

Taken together, these results suggest that EST data could be successfully mined to provide digital profiles of differential gene expression at the full genome level between normal and cancerous tissues. Our lists of transcriptional signatures might help to select candidate markers in cancer genetics or potential targets for therapy.

Increased expression of Bcl-xL in liver tumors

Bcl-2 family member Bcl-xL (Bcl2-associated X membrane protein) was selected for confirmation by immunoblotting due to its plausible biological role in cancer susceptibility. Moreover, both EST and SAGE results indicated that Bcl-xL was poorly expressed in normal liver, while abundant in other tissues (both normal and cancerous, data not shown), suggesting that this apoptotic regulator could constitute a good marker for liver cancer progression. As depicted in Figure 4, western blot analysis confirmed overexpression of Bcl-xL in a subset of human liver cancer specimens (hepatocellular carcinoma, adenocarcinoma but not cholangiocellular carcinoma) compared to normal liver (and placenta).

Identification of chromosome locations of differential gene expression in cancer

We next sought to analyze the chromosomal distribution of the genes which were over-expressed or repressed in tumor tissues. To this end, we mapped the previously identified genes showing significant differential expression between normal and tumor tissues along human chromosomes according to their banding, in order to build cancer-oriented transcriptome maps. To avoid possible biases due to chromosome length (e.g. chromosome Y as an obvious case) or different chromosomal gene densities, we computed the percentage of candidate genes against the total number of genes present on a particular chromosome or banding (see Table 2).

First, our results show that some chromosomes appear to be more active than others (Table 2A), with, for instance, chromosomes 15, 19 and Y being rarely involved in cancer-related gene expression changes compared to chromosomes 4 and 6. As expected from the results of Table 1, most chromosomal regions associated with changes in expression levels actually correspond to gene inactivation

Description	SKIN	BRAIN	BLOOD	MAMMARY GLAND	PLACENTA	LUNG	LIVER	TESTIS	EYE	COLON	BONE	MUSCLE	PROSTATE
CYTOKERATINS													
ENST00000330553 keratin 6L	4.01E-18												
ENST00000271963 keratin, type II, 1				3.34E-09									
ENST00000251643 Keratin, type I cytoskeletal 12 (Cytokeratin 12)									1.37E-08			3.95E-07	
ENST00000246635 Keratin, type I cytoskeletal 13 (Cytokeratin 13)		2.89E-11											
ENST00000167586 Keratin, type I cytoskeletal 14 (Cytokeratin 14)		1.39E-125		6.73E-08					1.32E-07				
ENST00000254043 Keratin, type I cytoskeletal 15 (Cytokeratin 15)	6.11E-09	4.20E-08		1.05E-08									
ENST00000301653 Keratin, type I cytoskeletal 16 (Cytokeratin 16)		4.76E-07											
ENST00000225901 Keratin, type I cytoskeletal 19 (Cytokeratin 19)		1.06E-10				1.18E-15	1.25E-17		1.17E-39				
ENST00000209718 Keratin, type I cytoskeletal 23 (Cytokeratin 23)				5.16E-10	5.13E-09								
ENST00000252242 Keratin, type II cytoskeletal 5 (Cytokeratin 5)		7.71E-94		1.77E-19									
ENST00000312899 Keratin, type II cytoskeletal 6A (Cytokeratin 6A)													1.23E-46
ENST00000252252 Keratin, type II cytoskeletal 6F (Cytokeratin 6F)				1.57E-10									
ENST00000293308 Keratin, type II cytoskeletal 8 (Cytokeratin 8)	2.02E-10			3.76E-34		1.53E-10	5.29E-33		1.80E-23				
ENST00000228652 KRT18 KERATIN TYPE I IPR001664/3054/2957				9.88E-38					2.47E-31				
ENST00000332088 KERATIN TYPE I IPR001664/2957/2952		2.89E-11											
COLLAGENS													
ENST00000225964 Collagen alpha 1(I) chain precursor				5.78E-12					3.23E-31				3.46E-81
ENST00000304636 Collagen alpha 1(III) chain precursor		2.33E-11		1.94E-13					2.08E-19				
ENST00000261037 Collagen alpha 1(VIII) chain precursor													
ENST00000297268 Collagen alpha 2(I) chain precursor	9.74E-09			1.53E-07				7.72E-09	1.12E-33			6.96E-08	
ENST00000295550 Collagen alpha 3(VI) chain precursor				3.57E-07									
ENST00000257309 Collagen alpha 2(IV) chain precursor				2.39E-09									
ENST00000300527 Collagen alpha 2(VI) chain precursor		4.76E-07											1.05E-19
ENST00000326158 Collagen alpha 3(X) chain precursor					8.39E-18								

Figure 1
Patterns of differential expression for collagen and cytokeratin genes in multiple normal and tumor tissues.
 The data are shown in a table format, in which rows represent individual genes and columns represent individual normal tissues. The color in each cell reflects the differential expression level of the corresponding gene in a particular tissue. A four color code was used to represent gene induction and repression in cancer libraries (dark green: 'normal-specific', i.e. not expressed in tumor libraries; light green: downregulated in tumor libraries; orange: upregulated in tumor libraries; red: 'tumor-specific'). If there was no significant change in gene expression between normal and tumor libraries or in case of missing/excluded data, the gene was given in a black color. The number inside the colored cells indicates the statistical significance (p -value < 0.01 after Bonferroni correction). See additional information for the full data.

patterns in cancer cells (373 up-regulated versus 744 down-regulated hits), striking examples of cancer-associated inactivation of gene expression being chromosome 17 and chromosome 3. While most tissues (14/16) were clearly subject to these cancer-associated gene inactivation patterns (especially lung, eye, colon, prostate and stomach), two tissues (tumor blood and liver) did not follow this trend. Chromosomal regions displaying at least five hits were further listed and this rough analysis was sufficient to detect 11 and 29 regions of clustering of up- and down-regulated genes, respectively (Table 2B). We found previously identified chromosomal regions associated with either tumor amplicon (e.g. 12q13) or deleted (e.g. 11p15) regions in tumors [23-26]. Interestingly, some of the chromosomal locations which were identified show tissue specificity, e.g. 12q13.3 in muscle. Moreover, in some cases, candidate genes could be contiguous or clustered in limited banding intervals. For instance, 19q13 is associated in tumor tissues of placental origin with complete extinction of eight clustered genes, namely pregnancy-specific beta-1-glycoproteins PSG-1, -2, -3, -4, -5, -6, -9 and -11. These genes belonging to the carcinoembryonic antigen family encode the major placental proteins found in maternal circulation during pregnancy [27,28].

In conclusion, in addition to providing differential expression profiles for individual genes, our EST-based procedure identified discrete regions on specific chromo-

somes that are enriched in genes deregulated in cancer libraries.

Discussion

Owing to advances in biotechnology and bioinformatics, researchers can now capture "molecular portraits" of various particular cancers using gene chips or SAGE data. These methods provide information on tens of thousands of genes simultaneously, and some variations in genes might be directly related to the cancer phenotype [1,29]. As multi-dimensional analysis of EST data is analogous to microarray experiments, we used the virtual differential display methodology to identify genes differentially expressed in normal versus tumor tissues. Our comprehensive approach gives an overview of numerous candidate genes which may be useful as improved biomarkers for diagnosis or as targets for developing novel treatment methods. For instance, EST-based formulation of collagen, integrin or cytokeratin expression profiles may have potential as a diagnostic aid for the detection of both tumor formation and development. Noteworthy, for discovery of tumor-associated molecules, it may be beneficial to use a combination of various digital differential display procedures and experimental data on gene expression. This is illustrated by the identification of prostate-specific Ets factor as a novel marker for breast cancer both computationally [8,30] (and this study) and experimentally [30-32].

Description	SKIN	BRAIN	BLOOD	MAMMARY GLAND	PLACENTA	LUNG	LIVER	TESTIS	BONE	MUSCLE	PROSTATE	STOMACH
ENST00000292899 Hemoglobin epsilon chain	1.26E-11	2.04E-14			3.93E-24	3.92E-14	3.73E-46			7.98E-09		
ENST00000261024 solute carrier family 40 (iron-regulated transporter), ferroportin 1					2.34E-07							
ENST00000274242 60S ribosomal protein L37	5.38E-48		8.55E-65									
ENST00000311111 60S ribosomal protein L38	3.35E-44		6.73E-29						7.79E-54			
ENST00000298289 60S ribosomal protein L44 (L36a)	2.46E-12		1.03E-09									
ENST00000257430 Adenomatous polyposis coli protein (APC protein)		5.86E-13						4.62E-09				
ENST00000227251 Alpha crystallin B chain	9.74E-09			2.77E-16	6.54E-08					3.53E-20		
ENST00000259396 Alpha-1-acid glycoprotein 1 precursor (AGP-1/OMD-1)						5.02E-09	3.10E-100					
ENST00000263574 Amyloid-like protein 2 precursor (CDEI-box binding protein)	3.57E-09				1.05E-16							
ENST00000300688 ATP synthase g chain, mitochondrial	8.57E-26		1.11E-09							7.66E-08		
ENST00000317296 Cohesin subunit SA-3		8.61E-08									3.27E-10	
ENST00000331242 cyclin D binding myb-like transcription factor 1		4.07E-08		1.69E-09								
ENST00000300177 cysteine knot superfamily 1, BMP antagonist 1; gremlin	3.63E-07					2.24E-07			5.50E-24			
ENST00000247655 Cytochrome c oxidase polypeptide VIic, mitochondrial precursor	1.44E-24									4.18E-13		
ENST00000265177 DC2 protein		2.59E-23				2.05E-09						
ENST00000228330 Decorin precursor (Bone proteoglycan II)					5.78E-12	2.05E-17	5.53E-07					
ENST00000256951 Epithelial membrane protein-1 (CL-20/B4B protein)				2.12E-09		1.36E-08						
ENST00000251595 Hemoglobin alpha chain	1.52E-20			2.24E-16	2.53E-112	2.10E-10	1.24E-35		2.36E-35	1.35E-07		
ENST00000009530 HLA class II histocompatibility antigen, gamma chain (CD74 antigen)				1.72E-28								4.71E-37
ENST00000319402 Ig alpha-1 chain C region				7.96E-34								9.07E-154
ENST00000259659 Integral membrane protein 2B					2.22E-09			5.71E-08				
ENST00000281455 Long-chain-fatty-acid-CoA ligase 2	6.95E-11						6.19E-10					
ENST00000266718 Lumican precursor					1.44E-35	9.79E-16						
ENST00000301522 Peroxiredoxin 2 (NKEF-B)	1.26E-14		5.06E-71									
ENST00000294390 phosphatidic acid phosphatase type 2B					4.84E-08						3.67E-29	
ENST00000275080 Putative mucin core protein 24 precursor (CD164)				1.33E-12	1.13E-08							
ENST00000258733 Putative transmembrane protein NMB precursor (glycoprotein HGFIN)				2.08E-07	3.79E-08							
ENST00000249786 small EDRK-rich factor 2	1.12E-62								3.78E-30			
ENST00000280356 Succinate dehydrogenase [ubiquinone] cytochrome B small subunit						6.34E-11	1.27E-18			1.66E-12		
ENST00000260356 Thrombospondin 1 precursor						3.00E-08			1.11E-27			
ENST00000330244 40S ribosomal protein S17					3.01E-40	5.05E-16						
ENST00000268664 40S ribosomal protein S2 (S4) (LLREP3 protein)				9.15E-46	7.02E-156	1.74E-94	6.37E-27					
ENST00000274065 40S ribosomal protein S3a					5.04E-86		4.19E-20					
ENST00000298087 54 kDa nuclear RNA- and DNA-binding protein (p54nrb)		5.56E-26			7.85E-26					1.75E-26		
ENST00000216146 60S ribosomal protein L3 (HIV-1 TAR RNA binding protein B)					3.61E-68			2.92E-31		6.82E-60		
ENST00000277425 60S ribosomal protein L7a (PLA-X polypeptide)				6.13E-18	5.38E-46							
ENST00000233154 Cytoplasmic protein NCK2	4.21E-10				6.31E-08							
ENST00000263367 E1B-55kDa-associated protein 5 isoform a	3.07E-09									3.98E-11		
ENST00000293831 Eukaryotic initiation factor 4A-1 (eIF4A-1)							3.91E-14			8.95E-14		
ENST00000316486 Fascin (Singed-like protein)		6.42E-19								2.07E-12		
ENST00000256442 G2/mitotic-specific cyclin B1					8.53E-18					8.99E-08		
ENST00000333261 Heterogeneous nuclear ribonucleoprotein A1				2.37E-07	1.82E-35			1.79E-23		4.13E-29		
ENST00000271095 Histone deacetylase 1 (HD1)				4.91E-17								1.08E-12
ENST00000326739 Inosine-5'-monophosphate dehydrogenase 2					1.44E-30					5.69E-09		
ENST00000330017 Interleukin-1 receptor-associated kinase 1 (IRAK-1)	1.33E-08	1.03E-07										
ENST00000296930 Nucleophosmin (NPM)					1.42E-33			9.09E-36				
ENST00000295688 T-complex protein 1, gamma subunit (TCP-1-gamma)										9.75E-18	1.12E-09	
ENST00000238272 WD repeat domain 34						2.32E-08	5.21E-07					
ENST00000241454 60S ribosomal protein L21						6.95E-13		3.81E-07				
ENST00000262584 60S ribosomal protein L8	6.37E-21					1.36E-18				2.75E-10		
ENST00000247550 BA528D8.2 (Novel protein similar to KIAA1074)	4.08E-15	9.33E-14		1.61E-07			9.05E-10					
ENST00000265056 DNA replication licensing factor MCM2 (Nuclear protein BM28)		1.25E-08			2.82E-22					7.47E-10		
ENST00000053867 Granulins precursor (Proepithelin)	3.02E-17											
ENST00000268027 Guanine nucleotide-binding protein G(i), alpha-2 subunit	1.57E-07			3.04E-07								
ENST00000311487 High mobility group protein HMGI/HMG-Y	4.24E-11	6.81E-09										
ENST00000269905 Lon protease homolog, mitochondrial precursor	6.56E-09				4.26E-43							
ENST00000188376 Phosphate carrier protein, mitochondrial precursor (PTP)						2.81E-51	2.10E-23	7.27E-13				8.00E-10
ENST00000055077 Replication factor C 40 kDa subunit (A1 40 kDa subunit)		5.08E-08				2.31E-11						
ENST00000328923 Serine hydroxymethyltransferase, mitochondrial precursor							1.33E-10			2.16E-08		
ENST00000303103 Small nuclear ribonucleoprotein associated proteins B and B' (snRNP-B)	1.61E-19					2.67E-27				2.16E-08		

Figure 2
Genes whose transcripts varied significantly and consistently in abundance in at least two different tissues.
 Thirty genes were selected in each class of differential expression (upregulated, downregulated, tumor-specific or absent in tumors). The results are shown for twelve tissues. The legend is the same as in Figure 1. See additional information for the full data.

General limitations of EST-based strategies, which have been abundantly discussed elsewhere [4,33,34], include poor sequencing depth of the libraries, uncertainty concerning the origin of the samples, and differences in library sizes. In addition, analysis of tumor-related differential expression patterns of individual transcripts may have specific drawbacks. For example, cancer cells often proliferate more rapidly than adjacent normal cells and it is possible that, in some cases, the observed changes in transcript abundance may reflect a response to increased proliferation rather than transformation *per se*. One

related problem is that many cell types are often pooled together during the preparation of EST libraries. Given that most cancers start as growths of single cells, the lack of cell-type specific libraries is a major limiting factor of the method. Lastly, the determined variations in transcript expression may not correlate with similar variations in the abundance of the encoded protein, highlighting the need to experimentally test the computer-based predictions either by western blotting or immunohistochemistry. Our validation result showing that Bcl-xL protein expression was markedly increased in hepatocellular carcinoma and

Tissue	Description	P value
BLOOD	ENST00000221209 BCL2/adenovirus E1B 19-kDa protein-interacting protein 3 like (NIP3L)	1,15E-43
	ENST00000296417 Histone H2A.z	2,41E-28
	ENST00000215071 26S proteasome non-ATPase regulatory subunit 8/S14	2,21E-22
	ENST00000310911 Neutrophil defensin precursor (HNP-1/2)	3,25E-36
BONE	ENST00000265131 Tenascin precursor	2,14E-08
	ENST00000222271 Cartilage oligomeric matrix protein precursor (COMP)	4,04E-14
	ENST00000287603 Prothymosin alpha	1,23E-09
	ENST0000044462 Proteasome subunit alpha type 4 (component C9)	2,73E-07
BRAIN	ENST00000286452 Neuronal kinesin heavy chain 1 (NKHC 5A)	3,29E-19
	ENST00000261918 Semaphorin 7A precursor (Semaphorin L/K1/CD108)	3,14E-53
	ENST00000255029 brevican isoform 1; chondroitin sulfate proteoglycan BEHAB	7,14E-41
	ENST00000322395 14-3-3 protein sigma	1,37E-17
EYE	ENST00000228811 Proline-rich protein 4 precursor (Lacrimal proline-rich protein)	2,50E-24
	ENST00000252725 ARP2/3 complex 41 kDa subunit (P41-ARC)	6,36E-09
	ENST00000233084 ATP-dependent helicase DDX1 (DEAD-box protein 1)	4,70E-22
	ENST00000271558 Retinoic acid-binding protein II, cellular (CRABP-II)	5,63E-17
LIVER	ENST00000311190 Serum amyloid A protein precursor (SAA)	2,58E-88
	ENST00000302053 Fibrinogen alpha/alpha-E chain precursor	5,52E-116
	ENST00000224237 Vimentin	3,32E-22
	ENST00000250383 Dehydrogenase/reductase SDR family member 2 (HEP27 protein)	1,93E-21
LUNG	ENST00000276297 Rho-GTPase-activating protein 7 (Dlc-1)	1,33E-18
	ENST00000263228 ubiquitin-conjugating enzyme UBC3B	2,05E-50
	ENST00000310118 26S proteasome non-ATPase regulatory subunit 2	1,01E-09
	ENST00000311218 NULL (<i>60S RIBOSOMAL L29 IPR002673 Ribosomal L29e protein</i>)	1,54E-07
LYMPH	ENST00000332976 Wolf-Hirschhorn syndrome candidate 1 protein isoform 1	4,48E-15
	ENST00000304187 Ig kappa chain V-III region CLL precursor (Rheumatoid factor)	6,01E-11
	ENST00000263737 Tetratricopeptide repeat protein 7 (TPR repeat protein 7)	5,61E-10
	ENST00000175506 Asparagine synthetase (TS11 cell cycle control protein)	7,31E-09
MAMMARY GLAND	ENST00000326227 Macrophage metalloelastase precursor (MMP-12)	1,08E-15
	ENST00000295522 Claudin-1	6,37E-21
	ENST00000323803 SAM pointed domain containing Ets transcription factor	1,23E-30
	ENST00000323636 MTO1 protein homolog (CGI-02)	6,75E-16
MUSCLE	ENST00000309889 Telethonin (Titin cap protein)	4,36E-37
	ENST00000291901 Troponin T	1,42E-82
	ENST00000250003 Myoblast determination protein 1 (Myogenic factor 3)	1,43E-22
	ENST00000303887 DNA replication licensing factor MCM7 (CDC47 homolog)	6,87E-18
PLACENTA	ENST00000222543 Tissue factor pathway inhibitor 2 precursor (Placental protein 5)	7,80E-101
	ENST00000301408 Choriogonadotropin beta chain precursor (CG-beta)	2,75E-21
	ENST00000268664 40S ribosomal protein S2 (S4) (LLREP3 protein)	7,02E-156
	ENST00000270172 DNA (cytosine-5)-methyltransferase 3-like	1,71E-15
PROSTATE	ENST00000310212 Tubulin tyrosine ligase-like protein HOTT1 (HQP0207)	1,23E-37
	ENST00000296755 Microtubule-associated protein 1B (MAP 1B)	3,33E-30
	ENST00000262030 ATP synthase beta chain, mitochondrial precursor	4,81E-07
	ENST00000315127 NULL (<i>IPR004443 YjeF-related protein</i>)	1,91E-15
SKIN	ENST00000289906 Apolipoprotein A-II precursor (Apo-AII) (ApoA-II)	5,73E-29
	ENST00000323496 thymosin-like 6	2,14E-66
	ENST00000329835 Melanocyte protein Pmel 17 precursor (ME20M/ME20S antigen)	5,10E-66
	ENST00000331803 Uridine phosphorylase 1	2,94E-09
STOMACH	ENST00000264498 Heparin-binding growth factor 2 precursor (HBGF-2/BFGF)	7,90E-28
	ENST00000238983 Triacylglycerol lipase, gastric precursor	3,12E-118
	ENST00000216281 Heat shock protein HSP 90-alpha (HSP 86)	1,57E-15
	ENST00000215631 Growth arrest and DNA-damage-inducible protein GADD45 beta	3,25E-15
TESTIS	ENST00000312511 Sperm protamine P1 (Cysteine-rich protamine)	1,15E-39
	ENST00000304434 homolog of yeast polyunsaturated fatty acid elongation enzyme 2	7,82E-15
	ENST00000311135 BLOCK 23	1,99E-33
	ENST00000322819 zinc finger, CCHC domain containing 7	4,41E-10

Figure 3
Genes whose transcripts exhibited tissue-specific differential expression in normal versus tumor libraries. This figure is a compilation of genes that appear to be differentially expressed in only one of the 15 studied tissues. The results are shown for fourteen tissues. The color code is the same as in Figure 1. See additional information for the full data.

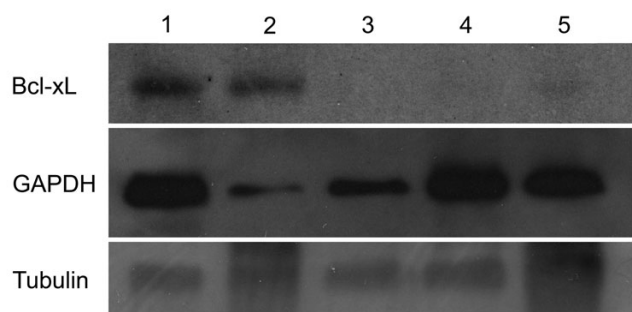


Figure 4
Western blot analysis of Bcl-xL expression in human normal and tumoral liver. Lane 1: hepatocellular carcinoma (male, age 65); lane 2: adenocarcinoma (male, age 52); lane 3: cholangiocellular carcinoma (male, age 46); lane 4: normal liver (male, age 24); lane 5: normal placenta (female, age 24). Bcl-xL immunoreactivity (26 kD) was observed in two out of three liver cancer samples (lanes 1–3). Normal samples (lane 4–5) had no signal for Bcl-xL expression. Note that GAPDH protein levels varied between normal tissues and cancer liver specimens and did not correlate with the mRNA levels predicted by the computer-based screen. The expression of tubulin was used as control for equal protein loading.

liver adenocarcinoma suggests that this Bcl-2 family member represent a potential marker for progression of a subset of liver cancers. Analysis of Bcl-xL immunoreactivity in more liver cancer specimens is needed to enhance the reliability of this finding. However, as it correlates with previous results [35-37], and in view of the pro-survival effect of Bcl-xL, we hypothesize that Bcl-xL overexpression could confer specific protection from death to several types of liver cancer cells compared to their healthy counterparts. If true, modulation of Bcl-xL expression level and/or activity might represent an interesting strategy to optimize the efficacy of chemotherapeutic agents in this particular tissue, as liver cancer represents a significant source of morbidity and mortality worldwide [38].

Aside from the proposal of potential diagnosis markers and targets for future cancer research, a more theoretical perspective of our study is the identification of critical factors that could influence differential gene expression levels in normal versus cancer cells, including genomic landscape features, e.g. levels of polymorphisms, chromosome breakpoints, gene density, GC content and chromatin methylation status. In this regard, although we cannot rule out the possibility of unidentified biases in our data mining procedure, our result showing a higher frequency of gene inactivation patterns in tumor tissues is intriguing, and sheds light on the importance of understanding the molecular mechanisms of negative gene regulation in cancer.

Conclusion

The final outcomes of the present work are identification of chromosomal regions frequently associated with aberrant expression in cancer libraries, description of differential expression profiles, and listing of cancer candidate genes (e.g. Bcl-xL) which may be useful as tissue-specific biomarkers for cancer diagnosis or as targets for anticancer research.

Methods

Data preparation

We have used an EST-based pipeline to scan for differential gene expression levels between normal and tumor states. Human ESTs from dbEST [39] (October 2004 release) were first extracted using the ACNUC sequence retrieval system [40]. ESTs were classified according to their UNIGENE library features [41] (October 2004). For each EST in dbEST, we extracted the accession code of the EST and the tissue or organ from which the EST library has been made. The tissue type was retrieved from the line containing 'Tissue_type', 'Tissue description', 'Organ' or 'Keyword'. This parsing approach stored no data when the tissue information did not appear in these fields, or in case of typographical errors or ambiguous aliases. ESTs that were labeled as coming from an unspecified tissue (e.g. 'mixed', 'pooled organs', 'cell line') or from a mixture of specified tissues, were discarded. The eVOC ontology [42] (October 2004) for anatomical sites and pathology types was then used to classify the libraries through a number of criteria such as tissue origin and pathological context including tumor state. This well-accepted hierarchical vocabulary provided us with a mean to determine when a specific tissue was part of an organ and when a specific label was part of the 'tumoral' state. A total of 5135 'tumor' and 2503 'normal' (i.e. non-pathological) libraries were catalogued. Our approach to EST clustering used the human genome as a reliable guide. ENSEMBL RNAs [43] annotated on human genome assembly (release 16.3) were used as a backbone for the clustering of dbEST sequences using MEGABLAST (alignment length = 100 bp and similarity = 95%) [44]. In order to avoid paralogous false positive assignment, only best EST hit matches were subsequently selected. RNA clustering of ESTs in both normal and tumor tissues was the starting point for digital differential analysis of gene expression.

Computer-based differential display procedure

The cDNA libraries were categorized into non-normalized or normalized/subtracted libraries by screening for the appropriate keywords in the original annotation of the respective dbEST entries (in the 'Keyword' and 'Library treatment' fields). All libraries for which none of the keywords were found were defined as being non-normalized. After removal of normalized and subtracted EST libraries, we created pools of equivalent EST libraries, i.e. libraries

Table 2: Chromosomal regions of differential gene expression in cancer. (A) Number of hits, i.e. number of genes with differential expression per chromosome, is depicted. 'Up' and 'down' mean chromosomal regions with increased and decreased tumor expression, respectively. '%' represents the percentage of candidate genes against the total number of genes present on the chromosome. (B) Chromosomal regions found to be associated with at least 5 hits in the digital subtraction analysis are shown ('banding'). '%' represents the percentage of hits for a particular chromosomal banding against the total number of genes present in the same banding. Chromosomal bandings marked in bold correspond to previously identified regions associated with either tumor amplicon ('Up' column) or deleted ('Down' column) regions in tumors.

Chromosome	A						B			
	Up		Down		Up+Down		Up		Down	
	Total Hits	%	Total Hits	%	Total Hits	%	Banding	%	Banding	%
1	33	1.02	54	1.68	87	2.70			1q21.3	3.12
2	11	0.75	28	1.90	39	2.65			1q32.1	3.20
									2p11.2	6.66
3	11	0.55	58	2.89	69	3.43			2p13.3	9.26
									2q35	4.55
4	38	2.26	46	2.73	84	4.99	3p21.31	4.03	3p21.31	3.14
5	4	0.67	8	1.35	12	2.02			4q13.3	9.37
6	14	1.22	36	3.15	50	4.37	6p21.31	6.33	5q33.1	6.58
7	11	1.05	19	1.81	30	2.86			7q21.3	8.93
8	18	1.25	30	2.08	48	3.34			7q22.1	3.32
9	31	1.69	39	2.13	70	3.82			9q34.11	3.60
10	5	1.05	10	2.10	15	3.15			9q34.3	3.60
11	30	1.32	46	2.03	76	3.36			11p15.1	7.81
12	25	1.14	61	2.78	86	3.91			11p15.4	2.56
									11p15.5	5.88
13	19	1.76	21	1.94	40	3.70			11q12.2	7.81
									12q13.13	4.80
14	5	1.27	8	2.03	13	3.29			12q13.3	10.12
									14q11.2	2.59
15	7	0.84	16	1.93	23	2.77			14q11.2	2.59
16	19	1.08	40	2.28	59	3.36			14q32.33	8.89
17	9	0.74	38	3.14	47	3.88	16p11.2	3.10	16p11.2	3.10
18	12	0.86	31	2.22	43	3.08			16q12.2	9.80
									17q21.2	5.71
19	14	0.80	29	1.65	43	2.44	17q25.3	2.81	17q21.2	5.00
20	10	0.56	38	2.13	48	2.69			17q23.3	8.06
21	8	0.70	25	2.18	33	2.87	19p13.3	2.43	19p13.11	2.15
22	17	1.20	31	2.19	48	3.40			19q13.2	3.69
X	21	1.59	31	2.34	52	3.93	21q22.3	3.12	19q13.31	7.04
Y	1	0.64	1	0.64	2	1.27			19q13.33	3.14
									20q13.12	2.92
									21q22.3	3.75
									Xq28	4.76

derived from the same tissue type and state (normal and tumor). Differential screening analysis was accomplished for a considered tissue when both normal and tumor pools displayed at least 10,000 ESTs. A total of 15 distinct paired tissue pools (blood, bone, brain, colon, eye, liver, lung, lymph, mammary gland, muscle, placenta, prostate, skin, stomach, testis) representing approximately 1.5 mil-

lion ESTs were therefore retained for the whole genome screening. Differential screening was performed for each tissue type individually using EST counts from tumors and corresponding normal counterparts. The relative expression of one particular gene in a tissue was characterized by the ratio of the number of ESTs matching this gene to the total number of ESTs sequenced in the respective tissue.

As such 'gene expression profiles' were derived from 'normal' and 'tumor' libraries, it was possible to build 2×2 contingency tables and then to apply the Fisher exact test against the null hypothesis that there was no association between a particular gene and the tumoral state. A p -value was determined for statistical significance and, because multiple tests were performed, a Bonferroni correction was applied on each pairs in order to reduce the false positive rate and to perform candidate gene sorting. Statistically significant hits showing at least 10-fold differences were compiled. Four classes of genes were defined, namely (i) genes displaying significantly higher expression levels in tumor tissues ('up-regulated' genes); (ii) genes displaying significantly lower expression levels in tumor tissues ('down-regulated' genes); (iii) genes expressed in tumor but not in normal tissues ('tumor-specific' genes); (iv) genes absent in the tumor types compared to normal tissues. Apart from the genes displaying absolute differences between normal and tumor condition, a ratio based on EST abundance in both conditions was computed to estimate the expression fold change for up- and down-regulated genes. Cytogenetic map position of the hits was inferred using ENSEMBL data (release 16.3). The pattern of expression of the differentially expressed transcripts ($n = 1190$, as determined by the EST analysis) in normal tissues was independently assessed by comparison to SAGE results obtained on the SAGE Genie website [45] and processed as previously described [46]. A total of 141 (non-tumoral) libraries containing more than 20,000 tags were partitioned into 19 normal tissues. The expression pattern of 13,435 transcripts was determined. Eight tissues (blood, brain, colon, liver, lung, mammary gland, placenta and prostate) were unambiguously mapped to the tissue terms used in the EST data mining procedure. From this sample, we queried as to which candidate transcripts associated with differential expression in a particular tissue (on the basis of the EST predictions) was expressed in the corresponding normal tissue (according to the SAGE data). Information on differential expression was also gained from reference to primary literature. As this effort corresponded to a manual task particularly unfitted to the large number of candidate genes presented here, we limited the analysis to the "up-regulated" and "down-regulated" lists related to the liver and breast tissues. We found that 54.2% (for liver) and 41.7% (for breast) of the annotated candidates identified through our computer-based screen were consistent with previously published data (see Additional files 1 and 3). The differential display procedure and other analytical steps were developed with R [47]. Expression and genomic data were stored in a local PostgreSQL database (GeMCore) [48] using PERL and Java script.

Western blot analysis

Nitrocellulose membrane was from Euromedex (Souffelweyersheim, France). The membrane was immunoblotted with anti-human Bcl-xL antibody (1:1 000 dilution, BD Pharmingen), and then with anti-mouse IgG antibody conjugated to horseradish peroxidase (1:5 000 dilution, Dako). Protein bands were revealed using enhanced chemiluminescence kit (ECL, Amersham). The membrane was stripped according to manufacturer's instructions and reprobed with anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) monoclonal antibody (1:1 000 dilution) and with anti-alpha-tubulin (1:1000 dilution, Sigma) to correct for differences in protein loading.

Authors' contributions

AA designed the study, performed the immunoblot assays, analyzed the data and drafted the manuscript. VN developed the algorithm for the differential display procedure, processed the SAGE data, participated in the data analyses and reviewed the manuscript. AB provided the antibodies and participated in the western blotting experiments. DM and CG provided funding and supervision for the work. All authors read and approved the final manuscript.

Additional material

Additional File 1

Upregulated candidates complete list. Upregulated genes in tumor tissues (complete list). Hits displaying at least a 10-fold increase in tumor-derived libraries compared to their normal tissue counterpart are shown. Chromosomal locations for each hit were inferred from Ensembl cytogenetic map. Hits were sorted by p value (exact Fisher's test; $p < 0.05$, Bonferroni corrected), ranked by expression ratio and ordered by tissue. Both known and novel ('NULL') transcripts are listed. 'Y': 'Yes'; 'ND': non-determined. Pubmed ID (PMID) is given for annotated candidate transcripts whose differential expression was documented in previously published data. '': in silico studies.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-94-S1.xls>]

Additional File 2

Downregulated candidates complete list. Downregulated genes in tumor tissues (complete list). Hits displaying at least a 10-fold decrease in tumor-derived libraries compared to their normal tissue counterpart are shown. Chromosomal locations for each hit were inferred from Ensembl cytogenetic map. Hits were sorted by p value (exact Fisher's test; $p < 0.05$, Bonferroni corrected), ranked by expression ratio and ordered by tissue. Both known and novel ('NULL') transcripts are listed. 'Y': 'Yes'; 'ND': non-determined. Pubmed ID (PMID) is given for annotated candidate transcripts whose differential expression was documented in previously published data. '': in silico studies.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-94-S2.xls>]

Additional File 3

Tumor specific candidates complete list. Complete list of genes absent from normal tissues and present in tumor types. Tumor-specific hits are shown. Chromosomal locations for each hit were inferred from Ensembl cytogenetic map. Hits were sorted by p value (exact Fisher's test; $p < 0.05$, Bonferroni corrected) and ranked by tissue origin. Both known and novel ('NULL') transcripts are listed. 'Y': 'Yes'; 'ND': non-determined.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-94-S3.xls>]

Additional File 4

Normal specific candidates complete list. Summary of genes absent from tumor types and present in normal tissues. Genes absent in the tumor types compared to normal tissues is shown. Chromosomal locations for each hit were inferred from Ensembl cytogenetic map. Hits were sorted by p value (exact Fisher's test; $p < 0.05$, Bonferroni corrected) and ranked by tissue origin. Both known and novel ('NULL') transcripts are listed. 'Y': 'Yes'; 'ND': non-determined.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-94-S4.xls>]

Additional File 5

Consistent candidates in multiple tissues. Genes whose transcripts varied significantly and consistently in abundance in at least two different tissues.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-94-S5.xls>]

Additional File 6

Tissue specific candidates. Genes whose transcripts exhibited tissue-specific differential expression in normal versus tumor libraries.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-94-S6.xls>]

Acknowledgements

VN is supported by a grant from INRA. AA is recipient of a fellowship from the ARC. The authors wish to thank Sandy Jacquier for critical reading of the manuscript, Dr. Pierre Colas at Aptanomics for providing the anti-Bcl-xL antibody and Dr. Cyrille Lamigeon for the anti-GAPDH antibody. We are grateful to Dr. Marie Semon for sharing the SAGE data.

References

- Gray JW, Collins C: **Genome changes and gene expression in human solid tumors.** *Carcinogenesis* 2000, **21(3)**:443-452.
- Rajkovic A, Yan MSC, Klysiak M, Matzuk M: **Discovery of germ cell-specific transcripts by expressed sequence tag database analysis.** *Fertil Steril* 2001, **76(3)**:550-554.
- Wang J, Liang P: **DigiNorthern, digital expression analysis of query genes based on ESTs.** *Bioinformatics* 2003, **19(5)**:653-654.
- Schmitt AO, Specht T, Beckmann G, Dahl E, Pilarsky CP, Hinzmann B, Rosenthal A: **Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues.** *Nucleic Acids Res* 1999, **27(21)**:4251-4260.
- Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K, Papadopoulos N, Vogelstein B, Kinzler KW, Strausberg RL, Riggins GJ: **A public database for gene expression in human cancers.** *Cancer Res* 1999, **59(21)**:5403-5407.
- Dahl E, Sadr-Nabavi A, Klopocki E, Betz B, Grube S, Kreutzfeld R, Himmelfarb M, An HX, Gelling S, Klamann I, Hinzmann B, Kristiansen G, Grutzmann R, Kuner R, Petschke B, Rhiem K, Wiechen K, Sers C, Wiestler O, Schneider A, Hoffer H, Nahrig J, Dietel M, Schafer R, Rosenthal A, Schmutzler R, Durst M, Meindl A, Niederacher D: **Systematic identification and molecular characterization of genes differentially expressed in breast and ovarian cancer.** *J Pathol* 2005, **205(1)**:21-28.
- Grutzmann R, Pilarsky C, Staub E, Schmitt AO, Foerder M, Specht T, Hinzmann B, Dahl E, Alldinger I, Rosenthal A, Ockert D, Saeger HD: **Systematic isolation of genes differentially expressed in normal and cancerous tissue of the pancreas.** *Pancreatol* 2003, **3(2)**:169-178.
- Scheurle D, DeYoung MP, Binninger DM, Page H, Jahanzeb M, Narayanan R: **Cancer gene discovery using digital differential display.** *Cancer Res* 2000, **60(15)**:4037-4043.
- Baranova AV, Lobashev AV, Ivanov DV, Krukovskaya LL, Yankovsky NK, Kozlov AP: **In silico screening for tumour-specific expressed sequences in human genome.** *FEBS Lett* 2001, **508(1)**:143-148.
- Brentani H, Caballero OL, Camargo AA, da Silva AM, da Silva WA Jr, Dias Neto E, Grivet M, Gruber A, Guimaraes PE, Hide W, Iseli C, Jongeneel CV, Kelso J, Nagai MA, Ojopi EP, Osorio EC, Reis EM, Riggins GJ, Simpson AJ, de Souza S, Stevenson BJ, Strausberg RL, Tajara EH, Verjovski-Almeida S, Acencio ML, Bengtson MH, Bettoni F, Bodmer WF, Briones MR, Camargo LP, Cavenee W, Cerutti JM, Coelho Andrade LE, Costa dos Santos PC, Ramos Costa MC, da Silva IT, Esteccio MR, Sa Ferreira K, Furnari FB, Faria M Jr, Galante PA, Guimaraes GS, Holanda AJ, Kimura ET, Leerkes MR, Lu X, Maciel RM, Martins EA, Massier KB, Melo AS, Mestriner CA, Miracca EC, Miranda LL, Nobrega FG, Oliveira PS, Paquola AC, Pandolfi JR, Campos Pardini MI, Passetti F, Quackenbush J, Schnabel B, Sogayar MC, Souza JE, Valentini SR, Zaiats AC, Amaral EJ, Arnaldi LA, de Araujo AG, de Bessa SA, Bicknell DC, Ribeiro de Camaro ME, Carraro DM, Carrer H, Carvalho AF, Colin C, Costa F, Curcio C, Guerreiro da Silva ID, Pereira da Silva N, Dellamano M, El-Dorri H, Espreafico EM, Scattone Ferreira AJ, Ayres Ferreira C, Fortes MA, Gama AH, Giannella-Neto D, Giannella ML, Giorgi RR, Goldman GH, Goldman MH, Hackel C, Ho PL, Kimura EM, Kowalski LP, Krieger JE, Leite LC, Lopes A, Luna AM, Mackay A, Mari SK, Marques AA, Martins WK, Montagnini A, Mourao Neto M, Nascimento AL, Neville AM, Nobrega MP, O'Hare MJ, Otsuka AY, Ruas de Melo AI, Paco-Larson ML, Guimaraes Pereira G, Pesquero JB, Pessoa JG, Rahal P, Rainho CA, Rodrigues V, Rogatto SR, Romano CM, Romeiro JG, Rossi BM, Rusticci M, Guerra de Sa R, Sant'Anna SC, Sarmazo ML, Silva TC, Soares FA, Sonati Mde F, de Freitas Sousa J, Queiroz D, Valente V, Vettore AL, Villanova FE, Zago MA, Zalberg H: **The generation and utilization of a cancer-oriented representation of the human transcriptome by using expressed sequence tags.** *Proc Natl Acad Sci U S A* 2003, **100(23)**:13418-13423.
- Aouacheria A, Navratil V, Wen W, Jiang M, Mouchiroud D, Gautier C, Gouy M, Zhang M: **In silico whole-genome scanning of cancer-associated nonsynonymous SNPs and molecular characterization of a dynein light chain tumour variant.** *Oncogene* 2005, **24(40)**:6133-6142.
- Du W, Lebowitz PF, Prendergast GC: **Elevation of alpha2(I) collagen, a suppressor of Ras transformation, is required for stable phenotypic reversion by farnesyltransferase inhibitors.** *Cancer Res* 1999, **59(9)**:2059-2063.
- Andreu T, Beckers T, Thoenes E, Hilgard P, von Melchner H: **Gene trapping identifies inhibitors of oncogenic transformation. The tissue inhibitor of metalloproteinases-3 (TIMP3) and collagen type I alpha2 (COL1A2) are epidermal growth factor-regulated growth repressors.** *J Biol Chem* 1998, **273(22)**:13848-13854.
- Moll R: **Cytokeratins in the histological diagnosis of malignant tumors.** *Int J Biol Markers* 1994, **9(2)**:63-69.
- Jiang Y, Harlocker SL, Molesh DA, Dillon DC, Stolk JA, Houghton RL, Repasky EA, Badaro R, Reed SG, Xu J: **Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays.** *Oncogene* 2002, **21(14)**:2270-2282.
- Holland EC, Sonenberg N, Pandolfi PP, Thomas G: **Signaling control of mRNA translation in cancer pathogenesis.** *Oncogene* 2004, **23(18)**:3138-3144.
- Ruggero D, Grisendi S, Piazza F, Rego E, Mari F, Rao PH, Cordon-Cardo C, Pandolfi PP: **Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification.** *Science* 2003, **299(5604)**:259-262.

18. Bader AG, Vogt PK: **An essential role for protein synthesis in oncogenic cellular transformation.** *Oncogene* 2004, **23(18)**:3145-3150.
19. Lai J, Flanagan J, Phillips WA, Chenevix-Trench G, Arnold J: **Analysis of the candidate 8p21 tumour suppressor, BNIP3L, in breast and ovarian cancer.** *Br J Cancer* 2003, **88(2)**:270-276.
20. Matsushima M, Fujiwara T, Takahashi E, Minaguchi T, Eguchi Y, Tsujimoto Y, Suzumori K, Nakamura Y: **Isolation, mapping, and functional analysis of a novel human cDNA (BNIP3L) encoding a protein homologous to human NIP3.** *Genes Chromosomes Cancer* 1998, **21(3)**:230-235.
21. Rountree MR, Bachman KE, Herman JG, Baylin SB: **DNA methylation, chromatin inheritance, and cancer.** *Oncogene* 2001, **20(24)**:3156-3165.
22. Robertson KD: **DNA methylation and chromatin – unravelling the tangled web.** *Oncogene* 2002, **21(35)**:5361-5379.
23. Wikman H, Nymark P, Vayrynen A, Jarmalaite S, Kallioniemi A, Salmenkivi K, Vainio-Siukola K, Husgafvel-Pursiainen K, Knuutila S, Wolf M, Anttila S: **CDK4 is a probable target gene in a novel amplicon at 12q13.3-q14.1 in lung cancer.** *Genes Chromosomes Cancer* 2005, **42(2)**:193-199.
24. Elkhouloun AG, Krizman DB, Wang Z, Hofmann TA, Roe B, Meltzer PS: **Transcript mapping in a 46-kb sequenced region at the core of 12q13.3 amplification in human cancers.** *Genomics* 1997, **42(2)**:295-301.
25. Moskaluk CA, Rumpel CA: **Allelic deletion in 11p15 is a common occurrence in esophageal and gastric adenocarcinoma.** *Cancer* 1998, **83(2)**:232-239.
26. Scelfo RA, Schwienbacher C, Veronese A, Gramantieri L, Bolondi L, Querzoli P, Nenci I, Calin GA, Angioni A, Barbanti-Brodano G, Negrini M: **Loss of methylation at chromosome 11p15.5 is common in human adult tumors.** *Oncogene* 2002, **21(16)**:2564-2572.
27. Hammarstrom S: **The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues.** *Semin Cancer Biol* 1999, **9(2)**:67-81.
28. Beckers JF, Zarrouk A, Batalha ES, Garbayo JM, Mester L, Szenci O: **Endocrinology of pregnancy: chorionic somatomammotropins and pregnancy-associated glycoproteins: review.** *Acta Vet Hung* 1998, **46(2)**:175-189.
29. Strausberg RL, Simpson AJ, Wooster R: **Sequence-based cancer genomics: progress, lessons and opportunities.** *Nat Rev Genet* 2003, **4(6)**:409-418.
30. Ghadersohi A, Sood AK: **Prostate epithelium-derived Ets transcription factor mRNA is overexpressed in human breast tumors and is a candidate breast tumor marker and a breast tumor antigen.** *Clin Cancer Res* 2001, **7(9)**:2731-2738.
31. Mitas M, Mikhitarian K, Hoover L, Lockett MA, Kelley L, Hill A, Gillanders WE, Cole DJ: **Prostate-Specific Ets (PSE) factor: a novel marker for detection of metastatic breast cancer in axillary lymph nodes.** *Br J Cancer* 2002, **86(6)**:899-904.
32. Katayama S, Nakayama T, Ito M, Naito S, Sekine I: **Expression of the ets-1 proto-oncogene in human breast carcinoma: differential expression with histological grading and growth pattern.** *Histol Histopathol* 2005, **20(1)**:119-126.
33. Imyanitov EN, Togo AV, Hanson KP: **Searching for cancer-associated gene polymorphisms: promises and obstacles.** *Cancer Lett* 2004, **204(1)**:3-14.
34. Qiu P, Wang L, Kostich M, Ding W, Simon JS, Greene JR: **Genome wide in silico SNP-tumor association analysis.** *BMC Cancer* 2004, **4(1)**:4.
35. Garcia EJ, Lawson D, Cotsonis G, Cohen C: **Hepatocellular carcinoma and markers of apoptosis (bcl-2, bax, bcl-x): prognostic significance.** *Appl Immunohistochem Mol Morphol* 2002, **10(3)**:210-217.
36. Takehara T, Liu X, Fujimoto J, Friedman SL, Takahashi H: **Expression and role of Bcl-xL in human hepatocellular carcinomas.** *Hepatology* 2001, **34(1)**:55-61.
37. Watanabe J, Kushihata F, Honda K, Sugita A, Tateishi N, Mominoki K, Matsuda S, Kobayashi N: **Prognostic significance of Bcl-xL in human hepatocellular carcinoma.** *Surgery* 2004, **135(6)**:604-612.
38. Guyton KZ, Kensler TW: **Prevention of liver cancer.** *Curr Oncol Rep* 2002, **4(6)**:464-470.
39. **DbEST: Expressed Sequence Tags database** [<http://www.ncbi.nlm.nih.gov/dbEST/>]
40. Gouy M, Gautier C, Attimonelli M, Lanave C, di Paola G: **ACNUC – a portable retrieval system for nucleic acid sequence databases: logical and physical designs and usage.** *Comput Appl Biosci* 1985, **1(3)**:167-172.
41. **Unigene: organized view of the transcriptome** [<ftp://ftp.ncbi.nih.gov/repository/UniGene/>]
42. **Evoque: expression ontology toolkit** [<http://www.eogenetics.com/evoke.html>]
43. **Ensembl database** [<http://www.ensembl.org/>]
44. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17)**:3389-3402.
45. Liang P: **SAGE Genie: a suite with panoramic view of gene expression.** *Proc Natl Acad Sci U S A* 2002, **99(18)**:11547-11548.
46. Semon M, Mouchiroud D, Duret L: **Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance.** *Hum Mol Genet* 2005, **14(3)**:421-427.
47. **The Comprehensive R Archive Network** [<http://stat.cmu.edu/R/CRAN/>]
48. **GeM (Genomic Mapping) Website** [http://pbil.univ-lyon1.fr/gem/gem_home.php]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

